


24-10-2011



INTERACCIÓN  
PERSONA-  
ORDENADOR

## TRABAJO DE DISPOSITIVOS DE INTERACCIÓN

Reconocimiento de voz | Adrián Neila Serrano

## Contenido

|  |   |
|--|---|
| 1. Nombres.....  | 2 |
| 1.1. Genéricos.....  | 2 |
| 1.2. Comerciales y libres. ....                                | 2 |
| 2. Tipo de interacción. ....                                   | 2 |
| 3. Paradigma de interacción. ....                              | 3 |
| 4. Estado comercial.....                                       | 3 |
| 5. Ventajas de su uso. ....                                    | 3 |
| 6. Posibles desventajas de su uso.....                         | 4 |
| 7. Interacción con otros dispositivos. ....                    | 4 |
| 8. Posibles usos del dispositivo. ....                         | 5 |
| 8.1. Dictado. ....   | 5 |
| 8.2. Control y comandos. ....                                  | 5 |
| 8.3. Telefonía.....  | 5 |
| 8.4. Médicos o destinados para personas con discapacidad. .... | 5 |
| 8.5. Aplicaciones integradas.....                              | 5 |

## 1. Nombres.

### 1.1. Genéricos.

Como nombre genéricos de dispositivos de interacción mediante reconocimiento de voz están: el conocido término de “manos libres”, el “ASR/RAH” (como acrónimo de automatic speech recognition/reconocimiento automático de voz) o el dictado por voz.

### 1.2. Comerciales y libres.

- EMBEDED VOICE de IBM.
  - XVOICE para linux.
- DRAGON NATURALLY SPEAKING de Nuance.
- SIRI de apple.
- CVOICECONTROL de Daniel Kiecza.
- OPEN MIND SPEECH (conocido formalmente como FreeSpeech) de Open Mind Initiative.
- SPHINX de CMU.
- ABBOT de SoftSound.

## 2. Tipo de interacción.

Los dispositivos de reconocimiento de voz tratándolos de manera particular son dispositivos de entrada. Pero debido a que nunca funcionan de manera aislada, ya que siempre interactúan con un computador u otro dispositivo se podría decir que son de entrada/salida ya que, claramente recogen información del exterior (voz) como parte de entrada y muestran una salida (a través de otros dispositivos) que puede ser desde un texto en una pantalla con lo que se ha dicho hasta la ejecución, de manera visible o transparente para el usuario, de la orden o comando que ha pronunciado.

### 3. Paradigma de interacción.

El paradigma de interacción actual de estos dispositivos sería el del tipo autómatas, ya que es un ordenador o dispositivo móvil el que se utiliza para lograr los objetivos de estos dispositivos. En un futuro, en un tiempo indeterminado, se podría considerar un paradigma de inteligencia ambiental, cuando se pueda hablar en cualquier sitio para interactuar con el entorno. Aunque también cabe la posibilidad de que el paradigma fuera el de las interfaces vestibles (wereables) debido a que es algo que va a ir, por definición de proximidad, unido a la herramienta de dicho dispositivo, la voz.

### 4. Estado comercial.

Los sistemas comerciales han estado disponibles ampliamente desde 1990 en su mayoría asociados a los computadores y más actualmente a los 'Smartphone'. Pero uno de los primeros dispositivos de reconocimiento de voz fue el [IBM SHOEBOX](#) allá por los años 1964 cuando se mostró en la Feria Mundial de Nueva York.

Hoy en día estos sistemas donde más está siendo más utilizado es en aplicaciones telefónicas, para dirigir a un usuario hasta el operador correspondiente al servicio que necesita: agencias de viajes, atención al cliente, información etc. La mejoría de estos sistemas de reconocimiento del habla han ido aumentando y su eficacia cada vez es mayor aunque distan mucho de entender la totalidad de la compleja comunicación humana hablada.

### 5. Ventajas de su uso.

La mayor ventaja de estos dispositivos es la que experimenta la gente con discapacidad, ya que se pueden beneficiar de los programas de reconocimiento de voz. Para las personas sordas o con problemas de audición el reconocimiento de voz se usa para generar subtítulos de conversaciones como en salas de conferencias, de lecturas o en servicios generales.

El reconocimiento de voz también es muy útil para las personas que tienen dificultades a la hora de usar las manos, que van por ejemplo desde lesiones leves por esfuerzos repetitivos (RSI) hasta discapacidades que impiden el uso convencional de los dispositivos de entrada de los computadores.

De hecho la gente que usa mucho el teclado se convirtió en un sector del mercado al alza para los dispositivos de reconocimiento de voz. También obtienen ventajas las personas con problemas de aprendizaje, las cuales tienen problemas en plasmar las comunicaciones en papel (ya que tienen mentalmente la idea correcta pero esta se vuelve incorrecta a la hora de plasmarla finalmente sobre un papel).

## **6. Posibles desventajas de su uso.**

La investigación (y la propia experiencia de los usuarios) muestra que en las ocasiones en las que, por problemas de reconocimiento del dispositivo, se pide replantear la pregunta o cuestión una y otra vez, provocan al usuario frustración y un estado de irritación, por lo que la mayores desventajas son las molestias debidas a falta de comodidad y precisión a la hora de que se entienda lo que los usuarios quieren expresar.

En los sistemas de reconocimiento de voz basados en entrenabilidad la problemática, por molesto y requerimiento de tiempo y esfuerzo, es el propio proceso de entrenamiento necesario para usar el dispositivo. La dependencia del hablante, de algunos dispositivos, también es un problema en el caso de que se requiera que el dispositivo sea usado por diferentes personas. Hay dispositivos que no soportan un habla continua reconociendo solo aquellas palabras que se diga entre pausas claramente definidas, por lo que es una desventaja ya que impide la supuesta mejora de velocidad que se pretende obtener y fuerza a unas acciones no naturales al ser humano. La mayoría de estos dispositivos requiere unas condiciones libre de ruidos u otras voces lo que supone un problema en todo el ámbito de la movilidad porque cada vez es más difícil encontrar espacios en silencio. Otra posible desventaja es la limitación del tamaño del dominio, ya que no se puede soportar todas las palabras del lenguaje de un idioma dado, aunque hay dispositivos de reconocimiento de voz que abarcan unas 200000 palabras, que es un rango bastante amplio.

## **7. Interacción con otros dispositivos.**

En la totalidad de los casos los dispositivos de reconocimiento de voz interactúan con un computador o con otro dispositivo, no funcionan de manera aislada ya que el dispositivo no es un servicio o un fin en concreto, sino una herramienta para interactuar con dichos elementos. Esto es, es un canal de comunicación entre el usuario y el ordenador.

## **8. Posibles usos del dispositivo.**

A pesar de que prácticamente cualquier tarea que comprenda la conexión con un computador potencialmente puede usar el reconocimiento automático del habla (RAH), las siguientes aplicaciones con las más frecuentes actualmente.

### **8.1. Dictado.**

El dictado es el uso más común para los sistemas de RAH en la actualidad. Esto incluye transcripciones médicas, dictados legales y comerciales, así como el procesamiento de textos en general. En algunos casos, es necesario usar terminología específica para vocabularios especiales para aumentar la precisión del sistema.

### **8.2. Control y comandos.**

Son sistemas de RAH que están diseñados para realizar funciones y acciones en un sistema operativo (normalmente), se comportan como sistemas de comandos y control. Usan expresiones como "Abrir Firefox" y "Iniciar un nuevo terminal" para hacer precisamente esa acción.

### **8.3. Telefonía.**

Algunos sistemas PBX (Private Branch Exchange, central privada) o de VozMail permite a quienes llaman usar comandos de voz en lugar de pulsar botones para enviar tonos específicos.

### **8.4. Médicos o destinados para personas con discapacidad.**

Muchas personas tienen dificultades para escribir, debido a limitaciones físicas, como lesiones por esfuerzo repetitivo (RSI), la distrofia muscular o muchos otros. Por ejemplo, las personas con problemas auditivos pueden usar un sistema conectado a su teléfono para convertir el discurso de la persona que llama en texto y poder así leerlo.

### **8.5. Aplicaciones integradas.**

Algunos de los teléfonos móviles más modernos incluyen reconocimiento de voz que permiten expresiones como "Llamar a casa". Esto podría ser un factor importante en el futuro del RAH. Como por ejemplo ¿Por qué no puedo hablar a mi televisión todavía?